**Business Analytics** é a prática de analisar dados para suportar a tomada de decisões.

É o estudo dos dados através de Data Science e de análise de operações, ou seja, análise de negócios. Isso compreende a formação de modelos preditivos, aplicação de técnicas de otimização e a comunicação dos resultados para os clientes, parceiros e executivos, com objetivo claro de resolver problemas de negócio.

Business Analytics exige a aplicação de métodos quantitativos, e baseados em evidências para modelagem de problemas de negócio e suporte a tomada de decisão.

Um dos componentes fundamentais em Business Analytics são os DADOS, de MUITOS DADOS. Para que a aplicação da ciência de dados mostre todo o seu valor.

Em Business Analytics aplicamos técnias de análise para resolver problemas de negócio. Portanto, Big Data tem uma participação fundamental dentro do conceito total de Business Analytics.

Business Analytics e Business Intelligence são a mesma coisa? Não.

Business Analytics e Data Science São a mesma coisa? Não. Data Science é um **CONJUNTO DE TÉCNICAS** que aplicamos em Business Analytics.

Business Analytic = Business + Big Data + Data Science Quando temos estes três componentes, é quando podemos aplicar o **Business Analytics.**



Cadeia onde se trabalha o Business Analytics

Dados brutos  
Fazemos agregações  
Extraímos inteligência destes dados  
Esta inteligência permite obter insights, informação que nem sequer cogitávamos   
Gestores tomam decisões a partir de insights gerados por analistas e cientistas  
Decisões causam impacto operacional na empresa  
Que geram resultado financeiros   
Para então obter Criação de Valor: o que importa de fato para a empresa

**Como criamos valor** a partir de tantos dados disponíveis?

Através da aplicação de BUSINESS ANALYTICS

O que é data mining? O que é mineração de dados?

Data mining é uma das técnicas usadas em Business Analytics. Quando falamos em trabalhar com Big Data, negócios e ciência de dados, nós temos na verdade uma série de TÉCNICAS à nossa disposição.

Uma dessas técnicas e **DATA MINING**. De forma simplificada a mineração de dados pode ser definida como um processo automático ou semiautomático. de explorar analiticamente grandes bases de dados com a finalidade de descobrir **padrões relevantes** que ocorrem nos dados e que sejam importantes para embasar a assimilação da informação, suportando assim, a geração de conhecimento.

Mineração de dados é o processo de exploração de grandes quantidades de dados com o objetivo de encontrar **anomalias**, **padrões e correlações** para suportar a **tomada de decisões** e proporcionar **vantagens estratégicas.**

Data mining é uma das térmicas de Business Analytics. Tornou-se ferramenta de apoio com papel fundamental na gestão de informação dentro das organizações. O volume de dados não para de crescer.

Algumas empresas passam a verificar a possibilidades de colocar os dados que elas geram e dominam como um **ativo da empresa** a ser considerado inclusive no balanço da empresa. Os dados são uma matéria-prima valiosa que quando analisados, geram **vantagem competitiva.**

Existe uma forte relação entre o volume de dados e o valor que eles podem gerar. Para gerar **resultados significativos** a mineração de dados exige um **processo constante de coleta, processamento e análise de informações.** Criando o Pipeline. Um fluxo contínuo de coleta, processamento e análise. As empresas estão cada vez mais criando esse tipo de pipeline.

Exemplos de Ferramentas de Data Mining



Ferramentas à esquerda são ferramentas visuais, utilizando o mouse para trabalhar com elas.

Ferramentas à direita requerem conhecimento em programação

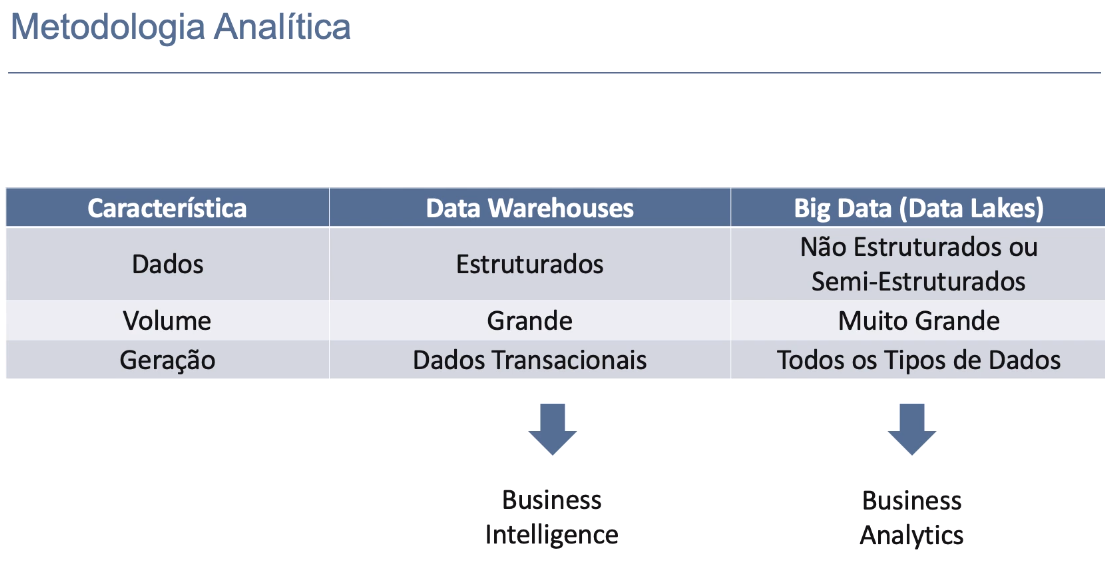
A metodologia ideal a ser usada em Business Analytics. Vamos começar analisando esta tabela.

Data Warehouse é um repositório de dados agregados, ou seja, para uma empresar operar no seu dia a dia, ela possui uma serie de banco de dados transacionais, um banco de dados para o sistema RP de gestão corporativa, banco de dados para CRM de gerenciamento de cliente, um banco de dados para página web da empresa e assim por diante.

No final do dia a empresa coleta dados desses bancos transacionais, faz uma agregação nos dados e coloca no **Data Warehouse.** No dia seguinte a equipe de Business Intelligence, equipe de análise de inteligência de negócios, extrai relatórios do Data Warehouse que então suportam a tomada de decisão.

O Data Warehouse foi concebido para trabalhar com **dados estruturados.** Os dados devem estar muito bem organizados para que eu possa trabalhar com Data Warehouse. Além disso um DW, tem um grande volume de dados e tem ainda **dados transacionais**, dados operações do dia a dia da empresa.

Hoje temos dados estruturados e semi-estruturados. Quando coletamos dados de redes sociais são dados não estruturados. Dados de mensagem de email, estes são semi ou não estruturados dependendo de como a **coleta** foi feita. Esse volume de dados desses dois últimos tipos é o que definimos como **Big Data.** E temos todos os tipos de dados.



Aplicamos Business Analytics para o Data Warehouse? Ou Business Intelligence para Big Data?

As técnicas de BI foram concebidas para trabalhar com **dados estruturados.** O Big Data é composto de dados não estruturados embora possa ter dados estruturados. Portanto Business Intelligence não foi concebido para finalidade de Big Data.

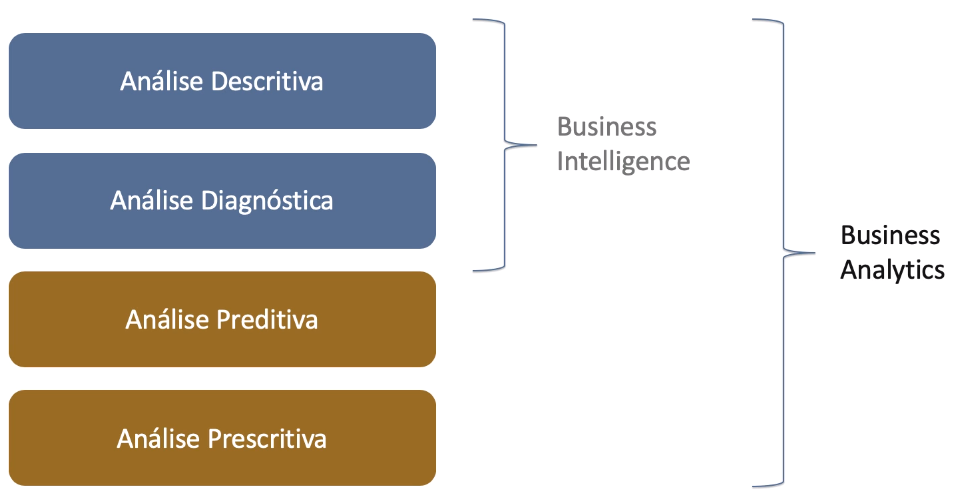
Por outro lado, as técnicas de Business Analytics que já trabalham com dados semi ou não estruturados podem ser usadas em dados estruturados.

Este diagrama mostra os quatro tipos principais de **análise** que nós temos nos dias de hoje.

Business Intelligence trabalha com os dois primeiros tipos de análise, a descritiva e diagnóstica. Enquanto o **Business Analytics** permite trabalhar com os quatro tipos de análises. Em geral trabalhamos com **análise preditiva e prescritiva.**

Análise descritiva e diagnóstica olham para o passado. Para o que já aconteceu, e a partir daí extrair informação que vai suportar a tomada de decisões futuras. O que é muito comum. Olhar para indicadores de vendas, produção, turnover de funcionários, tudo isso **já aconteceu.**

Hoje a massa de dados é tão grande, que nos permite **olhar para o futuro.** Com base em técnicas matemáticas e estatísticas nós conseguimos analisar grandes conjuntos de dados realizando análise preditiva e prescritiva. Prever o que pode acontecer e prescrever determinada ação de acordo com aquilo que os **modelos preditivos nos indicam.** Isso é possível devido ao grande volume de dados.



Na prática as técnicas de Business Intelligence têm foco em analisar o passado. Com base na análise do passado, tomadores de decisão podem prever o futuro. Neste caso não é a **técnica de análise** que está fazendo a previsão, mas sim o **ser humano.**

Ao trabalhar com Business **Analytics** estamos empregando técnicas que vão fazer as previsões por nós. Quando criarmos **modelos preditivos** para previsões, por exemplo, se um funcionário deixará ou não a empresa. O modelo fará uma **previsão.** O trabalho como cientista de dados é entregar um insight. Quem ainda vai **tomar a decisão é o ser humano.**

Os insights devem ser passados com muita responsabilidade. Vamos pegar dados do passado e então **alimentar o modelo,** e o modelo será capaz de prever se a pessoa deixará a empresa daqui um, dois, ou três anos. Permitindo uma gestão muito mais proativa.

O resultado de uma previsão é uma **estimativa** e a ciência que fundamenta esta metodologia é a **estatística.**

**Metodologia Analítica:**

Geração de dados: de onde estão vindo os dados que vão alimentar os modelos preditivos. Garantir que sejam dados verdadeiros, dados que representem a realidade. Os humanos são os responsáveis pela geração de dados, então temos que tomar os cuidados necessários. Em geral a responsabilidade do processo de geração de dados é do **Engenheiro de dados.**

Codificação dos dados: garantir que os dados tenham uma representação que possa ser usada nos modelos preditivos.

Captação de dados: Ao final das contas um pipeline deverá ser desenvolvido, uma coleta constante de dados. Coletar os dados, alimentar o modelo e fazer previsões. O processo de coleta também fica sob responsabilidade do **Engenheiro e dados**.

Transmissão de dados: talvez tenhamos que transferir dados entre sistemas diferentes.

Armazenagem de dados: temos que armazenar os dados com segurança, em um repositório que se possa ter acesso, de uma forma que mais tarde possamos usar os dados rapidamente e assim por diante.

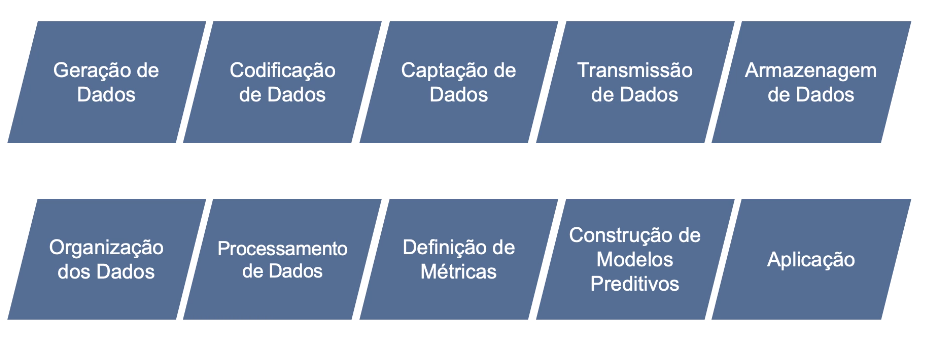
Organização dos dados: a responsabilidade a partir daqui muda. Ser o responsável por começar a dar uma estrutura aos dados. **Não existe análise de dados não estruturados.** Toda análise é em cima de dados estruturados. Dados não estruturados devem passar por um processo de estruturação e organização. Para trabalhar com texto é necessário **aplicar uma série de técnicas** para transformar esse texto em **dado estruturado.** Processamento de linguagem natural em Inteligência Artificial e Text Analytics.

Processamento de dados: realizar **análise exploratória, modificar variáveis, criar variáveis.** Faz parte do escopo do cientista de dados.

Definição de métricas: como saber se chegou ao **modelo preditivo ideal?** É necessário **definir as métricas antes de criar o modelo.** Se eu não definir as métricas, como vou saber que cheguei ao **modelo ideal?** Com 95% de acurácia. Se não conseguir **revisamos a métrica, revisamos o processo e revisamos os dados** para assim então dar início ao modelo.

Construção do modelo preditivo: cabe ao cientista de dados selecionar o melhor algoritmo de acordo com cada **problema.** Problema de regressão, problema de classificação, qual tipo de resultado entregar, qual volume de dados à minha disposição, qual o problema estou tentando resolver.

Aplicação: **usar o modelo para aquilo que ele foi construído.** Entregar insights, entregar previsões, entregar resultados para os tomadores de decisão. Dependendo do tipo de aplicação final, talvez o modelo tome a decisão por si. Na grande maioria dos casos passamos as previsões para que gestores possam decidir.



Realizar projetos de Business Analytics requer que se tenha um **objetivo claro e bem definido.** Todo projeto deve começar com uma **definição clara do problema de negócio** que se pretende resolver.

Desta forma podemos escolher quais **ferramentas, modelos preditivos, técnicas estatísticas** ou procedimentos deverão ser utilizados. Não existe uma única ferramenta que resolva todos os problemas e para uma escolha adequada o problema a ser resolvido deve estar bastante claro.

Isso significa economia de tempo, dinheiro, recursos, mão-de-obra para qualquer empresa. Todo profissional precisa dominar a **habilidade de identificar e definir problemas**, mas esta habilidade é muito pouco praticada e quase nunca ensinada.

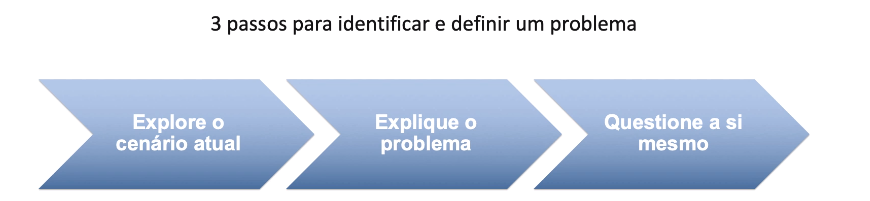
Identificar e definir um problema não é difícil, mas requer paciência, repetição e análise.

3 passos simples para identificar e definir um problema antes de pensar em ferramenta ou técnica a utilizar para a solução.

Explorar o cenário atual: é **fundamental,** o problema existe o hoje no presente, qual o impacto está sendo gerado? Quais as consequências? Quais são impactos e consequências em **NÃO RESOLVER** o problema? Quais são as possíveis causas? Qual a natureza do problema? É um problema inerente a uma área de negócio ou causado por pessoas? Isso ajudar no próximo passo.

Explique o problema: se você não sabe explicar o problema é porque ainda não o compreendeu. Você consegue compreender a essência do problema? Você é capaz de explicar o problema para outras pessoas?

Questione a si mesmo: Pergunte a si mesmo, **este é realmente o problema?** Você pode chegar à conclusão de que a resposta é simplesmente não. **O problema na verdade pode ser outro.** E na verdade o que se estava analisando **era a consequência do problema.** Todo esse processo requer repetição como um modelo de Machine Learning.



Este exercício do muito trabalho em se fazer e explorar estes três passos, por isso que **muita gente não faz.** O resultado é o que vemos, ninguém consegue definir um problema, ninguém sabe o que está acontecendo, a empresa é obrigada a contratar consultores.

Uma outra técnica que pode ajudar a **definir o problema de negócio** é utilizar o chamado **Toyota 5 Why’s 6 sigma –** Programa de gestão empresarial.

Pergunte cinco vezes por quê? Ao final teremos uma visão clara do problema que queremos resolver. Se a visão ainda não estiver clara, devemos estar de frente de um problema de proporções muito maiores. Talvez mirando na **consequência,** talvez não tenha feito as **perguntas de maneira correta**, talvez ainda **não tenha identificado o problema.**

**TUDO DEPENDE DESTA DEFINIÇÃO DE PROBLEMA.** Existem dezenas de linguagens de programação, diversos bancos de dados, diversos sistemas, ferramentas e técnicas. Cada qual mais apropriado para determinado tipo de problema.



A escolha da melhor solução depende da clara definição do problema.

Há uma tendência de queda nas buscas por Business Intelligence. Com Business Analytics nós conseguimos fazer muito mais coisas, além de termos o Big Data à nossa disposição.

Estudar Business Analytics é uma decisão inteligente me termos profissionais. Cada vez mais as pessoas vão buscar por esse tipo de conhecimento no mercado. Tanto as empresas como os profissionais.

Nós vamos focar principalmente em **análise preditiva e análise prescritiva, nós queremos olhar para o futuro.** Vamos usar **dados históricos**, preciso ainda de dados históricos para **alimentar um algoritmo** e a partir daí fazer **previsões.**

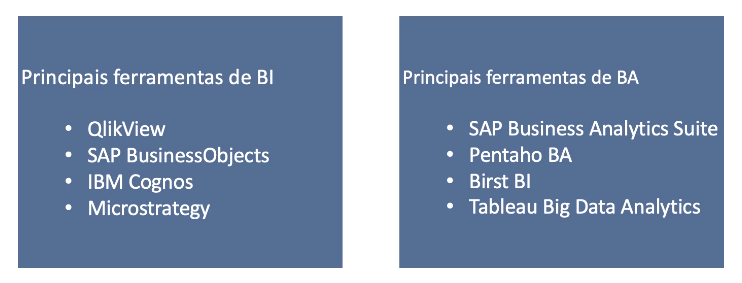
Com base nas vendas do passado, quanto podemos faturar no próximo ano?

Com base no volume de transação do cartão de crédito, qual a probabilidade de ter uma operação fraudulenta?

Com base na coleta de dados das redes sociais, qual a probabilidade de termos uma campanha de marketing trazendo aumento nas vendas de determinado produto?

Esta é a proposta de um treinamento em Business Analytics.

Para diferentes áreas de negócio, trazer diferentes problemas e aplicar Analytics. Técnicas que em geral utilizam **ciência de dados** e **data mining.**





Enquanto as aplicações estatísticas tradicionais se concentram em conjuntos de dados relativamente pequenos, a Ciência de Dados envolve **quantidades muito grandes**, normalmente o que chamamos de Big Data.

Quando falamos de ciência de dados, estamos falando em geral de aplicar **técnicas de análise** a grandes conjuntos de dados, que requerem também a aplicação correta de ciência de computação. Para **armazenamento** e **processamento de grandes conjuntos de dados.**

Com base nisso, **Business Analytics**, de um modo geral, deve ser visto como um **processo.** Um fluxo de atividades para que ao final se consiga resolver determinado problema. Para cada uma das etapas existem **técnicas, procedimentos, ferramentas** que podem ser empregadas.

Uma empresa que realmente deseja obter sucesso a partir da aplicação de Business Analytics devem compreender isso. Deve acima de tudo fomentar a construção de processo. Uma **linha de produção.** De acordo com os **dados** e com o **problema** técnicas diferentes serão utilizadas em cada etapa, mas **sabendo a sequência** de como tudo isso funciona, esse processo fica muito mais fácil de ser implementado.

São várias etapas interrelacionadas:

1.Definiação clara do problema de negócio a ser resolvido: Este é o **ponto de partida.** A definição clara do problema de negócio é que vai definir **tudo que será feito daqui em diante.** Não definir claramente o problema, é a certeza do processo de Analytics mal sucedido.

2.**Armazenamento de dados** eficiente e os passos de **pré-processamento** de dados, pois estes são críticos para o sucesso da análise.

3.É preciso selecionaras **variáveis de resposta apropriadas** e decidir sobre o número de **variáveis a serem investigadas**. A variável é como se fosse a coluna, e as linhas fossem observações. Quando coletamos dados da fonte para resolver um problema, nem todas as variáveis são relevantes. Portanto é necessário aplicar outro filtro de **técnicas** **para filtrar variáveis.** Uma das fases mais cruciais de todo o processo, a escolha errada das variáveis pode **comprometer** todas as etapas seguintes.

4.Os dados precisam ser rastreados em busca de outliers e os valores faltantes (missing) precisam ser endereçados (com valores ausentes omitidos ou apropriadamente através de um dos vários métodos disponíveis). Depois da identificação das variáveis importantes para o problema, preciso identificar **valores que fogem do padrão, valores que fogem da média dos dados.** Talvez, este seja exatamente o objetivo na prevenção de fraudes. Dependendo do problema é o **outlier que queremos identificar,** outros casos removê-lo para não influenciar na média dos dados e consequentemente **influenciar o modelo preditivo.**

5.Antes de aplicar modelos preditivos e métodos sofisticados, os **dados precisam ser visualizados e resumidos**. Negligenciar a visualização impede de identificar padrões ou problemas que não foram detectados durante a codificação. A **construção de gráficos** faz parte do processo de análise, ajuda a **compreender visualmente como os dados estão organizados.** A Análise Exploratória.

6.**Resumo dos dados** envolve estatísticas típicas de resumo como média, percentis e mediana, desvio padrão e correlação, bem como resumos mais avançados, tais como **componentes principais.** Aqui é onde os **dados são resumidos.**

7.Métodos apropriados de **modelagem preditiva** **precisam ser aplicados.** Dependendo do problema, isso pode envolver **regressão linear, regressão logística, árvores de regressão / classificação,** vizinho mais próximo, clustering, entre outros. É aqui onde temos os dados praticamente prontos e processados. A partir daqui **aplicar o algoritmo** para trabalhar com modelagem preditiva, fazer previsões.

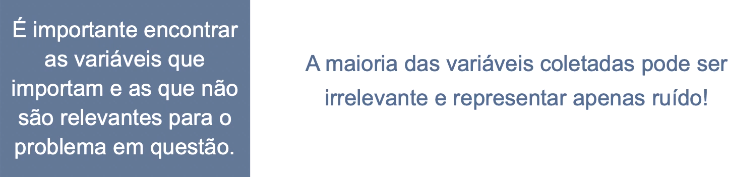
8.Finalmente, os insights da análise precisam ser implementados. É preciso **agir sobre os resultados.** Planejar, fazer, verificar e agir. O cientista de dados deverá trabalhar em conjunto com a área de negócio, explicar os insights obtidos através do modelo preditivo. Explicar como chegou a essa conclusão e propor alternativas. A decisão final é do diretor. **Fornecer suporte para área de decisão.**



Variáveis também são chamadas de atributos ou características. Observações também são chamadas de instâncias, registros ou mesmo casos. As **observações representam a quantidade de dados,** enquanto as **variáveis representam o formato dos dados.** Para um determinado problema de negócio, coletaremos dados, com muitas observações e com muitas variáveis.

No início do processo de análise não sabemos quais as variáveis mais importantes e isso será descoberto durante o processo, que aliás é uma etapa crítica. Será que todas as variáveis são importantes para tal processo de análise?

Encontrar duas variáveis **altamente correlacionadas.** Isso pode causar problemas principalmente em **tarefas de regressão.** Por conta dissotalvez tenhamos que excluir uma das variáveis



O fato de ter muitas variáveis não significa que isso é positivo. Um conjunto de dados com muitas dimensões pode causar uma serie de transtornos durante as etapas de análise. Precisamos **identificar as melhores variáveis** ou **reduzir a dimensionalidade.**

A mineração de dados explora e análise grandes quantidades de dados para **descobrir padrões significativos.** A escala de uma aplicação típica de mineração de dados, com grandes números de casos e muitas variáveis, excede a de uma **investigação estatística padrão**

A mineração de dados é uma técnica aplicada durante o processo de Business Analytics para **buscar padrões que ajudam a decidir quais variáveis manter** no conjunto de dados e quais remover, ou eventualmente quais variáveis corrigir, como outliers, missing ou qualquer outro tipo de problema ligado a ruído. Mineração de dados é fazer a **escolha ideal das variáveis.**

Mineração de dados são como estatísticas em escala e velocidade. A grande escala e os requisitos de velocidade dos dados criam uma grande demando por **automação**. A Mineração de dados utiliza uma combinação de **regras de conhecimentos de padrões**, que utilizam regras estatísticas, bem como regras extraídas através de **aprendizado de máquinas.**

Dentro da mineração de dados eu tenho um **conjunto de regras para buscar os padrões.** A partir daí eu consigo tomar decisões sobre a escolha das variáveis ou sobre qualquer outra decisão que desejada dento do conjunto de dados analisado.

A mineração de dados tem **ampla aplicabilidade,** com aplicações em **inteligência e análise de segurança, genética, ciências sociais e naturais e negócios** e muitos outros. Estudar quais compradores tem maior probabilidade de comprar, responder um anúncio, declarar falência, cometer fraudes ou abandonar serviços por assinatura são de **vital importância** para qualquer negócio. Muitos problemas de mineração de dados lidam com problemas de **resultados categóricos**, como apenas um sim ou não. Isso é o que torna os **métodos de aprendizado de máquina,** que têm suas origens na **análise de dados categóricos tão úteis.**

A estatística por outro lado tem suas origens nas **análises de dados contínuas**. Isso torna as estatísticas especialmente uteis para análise do **tipo de correlação.** Onde se passa através de um grande número de correlações para então encontrar aquelas com maior **nível de relacionamento.**

Devido às quantidades muitas vezes enorme de dados o papel dos conceitos estatístico tradicionais como: intervalo de confiança, teste de significância e outros, acaba sendo reduzido. Com grandes conjuntos de dados quase qualquer pequena diferença torna-se significativa. Isso é comum em modelos preditivos, que chamamos de **overfitting** quando o **modelo aprende demais a partir do conjunto de dados.**

Com conjuntos de dados cada vez maiores nós conseguimos resultados cada vez mais precisos. Mas para que isso aconteça precisamos da **combinação ideal das variáveis**, isso é quase uma ciência por si só. Usamos a mineração de dados como uma forma de buscar padrões, detectá-los e então fazer escolhas a partir disso. Nem sempre a estatística tradicional vai dar as respostas que precisamos e nisso podemos usar **Machine Learning para mineração de dados,** para assim escolher as melhores variáveis para alimentar outro modelo de **Machine Learning.**

Machine Learning é “apenas” uma ferramenta. Mineração de dados emprega Machine Learning, para escolhermos as melhores variáveis, continuarmos com o processo de Business Analytics e depois aplicar **outro modelo para fazer classificação, regressão** e assim por diante.